On Blind Separation of Complex-Valued Sources by Extended Hebbian Learning

Simone Fiori

***Abstract*—The aim of this letter is to present a nonlinear extension to Sanger's generalized Hebbian learning algorithm for complex-valued data neural processing, which allows for separating mixed independent circular source signals. The proposed generalization relies on an interesting interpretation of nonclassical Hebbian learning proposed by Sudjianto and Hassoun for real-valued neural units.**

***Index Terms*—Blind source separation, neural networks, Hebbian learning.**

I. Introduction

BLIND source separation (BSS) by extensions of principal component analysis (PCA) has raised much interest in the neural network and signal processing communities (see for example [5], [9] and references therein). In fact, it has been proven by some papers that adding nonlinearity to linear PCA neural networks makes them able to improve the independence of their outputs so as to allow blind separation of real-valued independent sources [5], [9]. Recently, some attempts have been made in order to extend the best known PCA algorithms to the complex case.

In this letter, we formally derive a new learning algorithm as a nonlinear complex extension of generalized Hebbian algorithm [10] for a linear feedforward network, the extended Hebbian learning algorithm (EHA). Also, in opposition to the commonly-adopted heuristic approaches for selecting the involved nonlinearity, we propose a possible choice under the theoretical framework of Sudjianto and Hassoun [11]. Then we show how a particular nonlinearity, the Rayleigh distribution, allows the neural network to perform blind separation of complex-valued circular source signals, which has several signal processing applications (see e.g., [1], [2], [4], [6]).

As in some recent research works [1], [6], the only information carried by the mixed signals that is exploited in this work is the one about the modulus, which indeed suffice to extract the hidden sources. This means that the proposed algorithm is phase-blind.

In order to numerically assess the behavior of the EHA method, computer simulation results are presented and discussed on parallel and sequential (cascade) version of the EHA algorithm.

Manuscript received November 21, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. G. Ramponi.

The author is with the Neural Networks and Adaptive System Research Group, Department of Industrial Engineering, University of Perugia, Italy (e-mail: sfr@unipg.it).

Publisher Item Identifier S 1070-9908(01)06593-2.

II. Extended Hebbian Learning Algorithm (EHA)

We consider a complex-weighted single-layer neural network, formed by linear units, which performs nonclassical PCA of complex-valued data. The network is described by the input vector $\mathbf{x} \in \mathcal{C}^p$, a set of weight-vectors $\mathbf{w}_k$ and outputs $y_k \stackrel{\text{def}}{=} \mathbf{w}_k^H \mathbf{x}$, where "$\cdot^H$" denotes conjugate transpose. The number of neurons of the network is denoted here with $m \leq p$.

Let the following criterion be defined

$$J(\mathbf{w}_k) \stackrel{\text{def}}{=} U(\mathbf{w}_k) + L(\mathbf{w}_k). \tag{1}$$

The criterion $U(\cdot)$ contains a nonlinear function of the kth neuron's output and is defined as follows:

$$U(\mathbf{w}_k) \stackrel{\text{def}}{=} E_{\mathbf{x}}[f(\mathbf{w}_k^H \mathbf{x})|\mathbf{w}_k] \tag{2}$$

where the symbol $E_{\mathbf{x}}[f|\mathbf{w}]$ denotes conditional expectation of f with respect to $\mathbf{x}$ subject to the hypothesis $\mathbf{w}$, hereafter simply written in short notation as $E[f]$. The $f(\cdot)$ is a real-valued, positive function of complex-valued argument that, by definition, is supposed of the form $f(z) \stackrel{\text{def}}{=} g(|z|)$, $z \in \mathcal{C}$, $g: \mathcal{R}_0^+ \to \mathcal{R}_0^+$, with $g(u)$ being continuously differentiable almost everywhere, nondecreasing with a unique minimum in $u = 0$.

As far as PCA is concerned, the adjoint function $L(\cdot)$ is used for adding to the criterion (1) the necessary constraints of orthonormality of the weight vectors, namely, $\mathbf{w}_\alpha^H \mathbf{w}_\beta = 0$ if $\alpha \neq \beta$ and $\mathbf{w}_\alpha^H \mathbf{w}_\alpha = 1$. The orthogonality conditions can be rewritten more conveniently by observing that $\mathbf{w}_\alpha^H \mathbf{w}_\beta = 0$, if and only if $\mathrm{Re}\{\mathbf{w}_\alpha^H \mathbf{w}_\beta\} = 0$ and $\mathrm{Im}\{\mathbf{w}_\alpha^H \mathbf{w}_\beta\} = 0$. Thus, the function $L(\cdot)$ may be expressed as

$$L(\mathbf{w}_k) \stackrel{\text{def}}{=} \sigma_{kk}(\mathbf{w}_k^H \mathbf{w}_k - 1) + \sum_{j=1}^{k-1} \mathrm{Re}\{\sigma_{kj}^\star \mathbf{w}_k^H \mathbf{w}_j\} \tag{3}$$

where a set of complex Lagrange multipliers $\{\sigma_{kj}\}$ has been introduced, and "$\cdot^\star$" denotes complex conjugation.

To search for optimal weights $\mathbf{w}_k^{\text{opt}}$ maximizing the criterion (1), a gradient steepest ascent (GSA) learning algorithm is employed here. The optimal multipliers as functions of the $\mathbf{w}_k$s can be found by solving equations $\mathbf{w}_h^H(\partial J/\partial \mathbf{w}_k) = 0$ for different values of the index h. After straightforward calculations (whose real-valued counterpart may be found in [9]) by defining

$$\mathbf{P}_k \stackrel{\text{def}}{=} \mathbf{I} - \sum_{j=1}^{k} \mathbf{w}_j \mathbf{w}_j^H,$$
$$G(\zeta) \stackrel{\text{def}}{=} \frac{dg(|\zeta|)}{d|\zeta|}\,\frac{1}{|\zeta|} \qquad \text{with } \zeta \in \mathcal{C} \tag{4}$$

the new complex nonclassic counterpart of GHA learning rule (or EHA) writes

$$\frac{d\mathbf{w}_k}{dt} = \eta \mathbf{P}_k E[G(y_k) y_k^\star \mathbf{x}], \qquad k = 1, 2, \ldots, m \tag{5}$$

with η being a positive learning step-size. The factor $E[G(y_k)y_k^\star \mathbf{x}]$ may be interpreted as a complex nonclassical Hebbian term common to each neuron, while projector $\mathbf{P}_k$ is a deflating factor that orthogonalizes the weight-vector $\mathbf{w}_k$ with respect to the others in the network.

III. Application to Blind Separation

Following some studies on nonlinear and robust principal component analysis, in [11], Sudjianto and Hassoun considered the problem of learning by optimization of the nonquadratic criterion $J(\mathbf{w}) \stackrel{\text{def}}{=} E[S^2(\mathbf{w}^T\mathbf{x})]$ subject to the restriction $\mathbf{w}^T\mathbf{w} = 1$, where $y = \mathbf{w}^T\mathbf{x}$ is the output of a single-unit real-weighted neural network, and $S(\cdot)$ is a generic saturating sigmoidal function, such that $S(\cdot) \in [-1, +1]$. Sudjianto and Hassoun experimentally proved that a two-input/one-output neuron, following a two-input/two-output PCA-based prewhitening network, is able to separate a signal from a linear mixture of two signals (e.g., a sinusoid corrupted by Gaussian nose), provided that the nonlinearity $S(\cdot)$ coincides with the cumulative distribution function of the unwanted signal (e.g., of the noise). This principle was recently extended by the present author [8] to a three-layer network for performing information-theoretic-based independent component analysis. However, the existing networks based on it cannot deal with complex-valued sources, owing to some of their inherent structural/learning limitations.

Let us consider now the extension of the previous theory to the complex case. We define the cost function $U(\mathbf{w}) \stackrel{\text{def}}{=} E[S^2(|y|)]$ for a complex-weighted neuron with output $y = \mathbf{w}^H\mathbf{x}$. Its GSA maximization under the constraint $\mathbf{w}^H\mathbf{w} = 1$ yields the learning rule

$$\frac{d\mathbf{w}}{dt} = \eta(\mathbf{I} - \mathbf{w}\mathbf{w}^H) E\left[\ell(|y|)\frac{y^\star}{|y|}\mathbf{x}\right] \tag{6}$$

with $\ell(u) \stackrel{\text{def}}{=} 2S(u)S'(u)$. It closely recalls (5) for $k = 1$. In our case, we assume $S(|y|) = Q_{|Y|}(|y|) \stackrel{\text{def}}{=} \int_0^{|y|} q(u)\, du$, where $q(\cdot)$ represents the discriminant probability density function. By equating (6) with (5), $k = 1$, it is possible to find the relationship between $q(\cdot)$ and $g'(\cdot)$, which writes $g'(u) = 2q(u)\int_0^u q(u)\, du$.

Concerning the EHA rule, it is worth noting the importance of the deflating projectors $\mathbf{P}_k$, which cause each neuron to filter a different independent source in blind separation.

This principle can be employed for separating out independent complex-valued circular signals from their linear mixtures. Let us suppose input $\mathbf{x}$ contains a complex linear mixture of statistically independent signals [2], and that one of these signals is a Gaussian noise of the form $n = r + js$, where both r and s are zero-mean Gaussian random variables of variance σ^2. Then it is known that the modulus $|n|$ follows the Rayleigh distribution

$$q_{\mathrm{R}}(|n|) = \frac{|n|}{\sigma^2}\exp\left(-\frac{|n|^2}{2\sigma^2}\right).$$

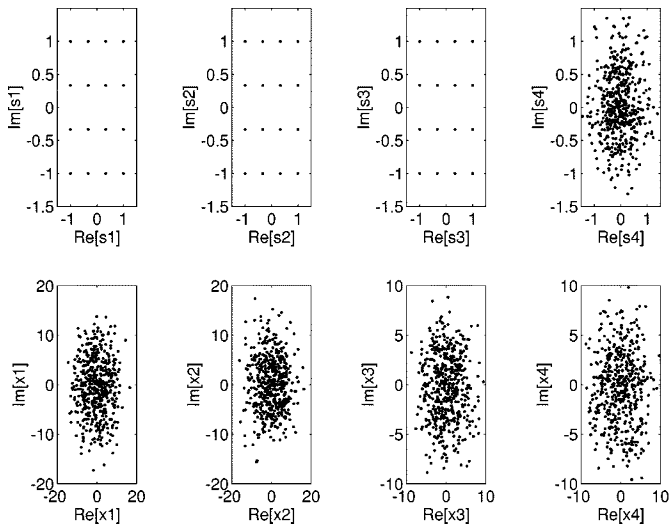

Fig. 1. Four independent signals and four mixtures of them.

Then by the latter formula mentioned earlier, we find

$$g_{\mathrm{R}}'(u) = \frac{2u}{\sigma^2}\left[\exp\left(-\frac{u^2}{2\sigma^2}\right) - \exp\left(-\frac{u^2}{\sigma^2}\right)\right]\Gamma(u)$$

where $\Gamma(u)$ is the unit-step function. By assuming in (5) the function $G(y_k)$ as the quantity $(g_{\mathrm{R}}'(|y_k|)/|y_k|)$, it is then possible to separate independent complex-valued signals mixed by a unitary operator. The general problem where generic linear mixtures are dealt with can be solved by prewhitening the data through *ad hoc* algorithms [4], [2].

IV. Computer Simulation Results

In order to test the explained blind separation algorithm, computer simulations inspired by [2] are showed and discussed in the following.

A. Four Mixed Source Signals

As a numerical example, suppose input $\mathbf{x} \in \mathcal{C}^4$ is formed by a linear mixture of four independent signals arranged in a vector $\mathbf{s} \in \mathcal{C}^4$. Signals s_1 to s_3 are QAM16; signal s_4 is a Gaussian noise. The mixture is computed as $\mathbf{x} = \mathbf{M}\mathbf{s}$, where $\mathbf{M}$ is a randomly generated 4×4 complex matrix. The first row of Fig. 1 depicts the independent signals while second row shows the obtained four mixtures.

By means of the Sudjianto–Hassoun principle, a linear neural network with four inputs and four outputs, trained by the learning rule (5) with the Rayleigh nonlinearity, should be able to recover the independent signals except for a phase shift and a random permutation [4], after mixture prewhitening. Simulation results are shown in Fig. 2. The first row depicts the result of prewhitening performed by means of the well-known Laheld–Cardoso's standardizing algorithm [2], and the second row shows the last 100 outputs of the network trained by (5) on the prewhitened data. Fig. 3 depicts the absolute values of source-to-output separation matrix $\mathbf{\Pi}$ defined by $\mathbf{y} = \mathbf{\Pi}\mathbf{s}$. At convergence, it should be of the form $\mathbf{\Pi} = \mathbf{P}\mathbf{D}$, where $\mathbf{P}$ is a permutation and $\mathbf{D}$ is complex diagonal, both being arbitrary

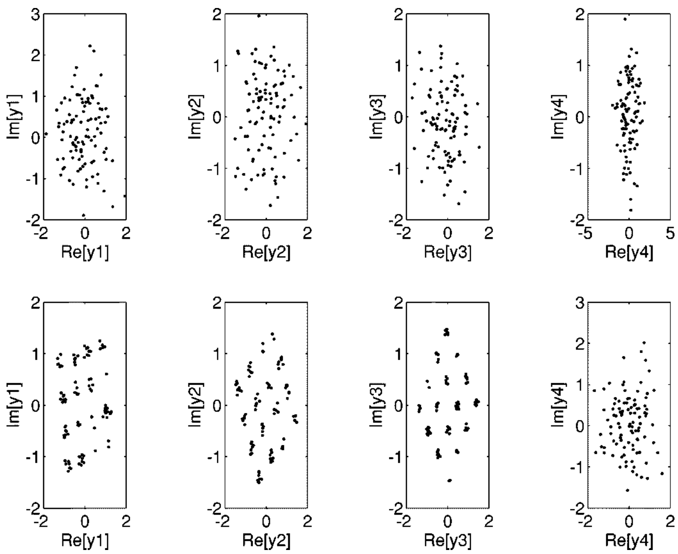

Fig. 2. Standardized mixtures and network's output after learning by rule (5).

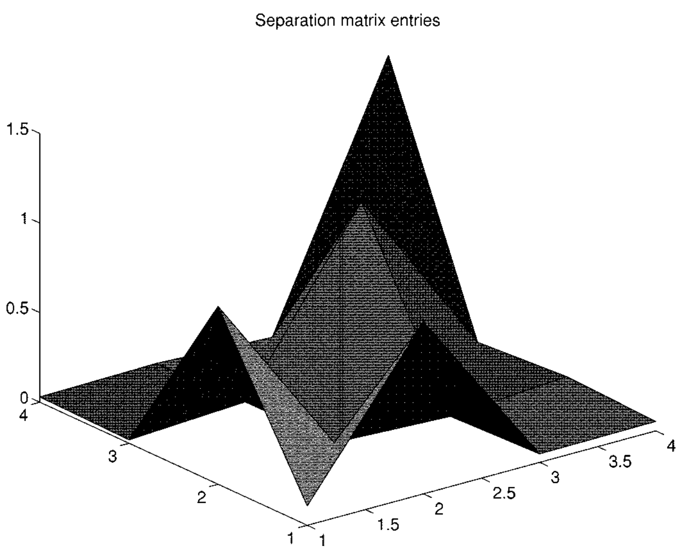

Fig. 3. Source-to-output separation matrix.

[4]. As expected, only one entry per row is approximately different from zero.

Successful simulation results show that the network is able to recover the independent signals.

B. Parallel versus Sequential Operation

Together with the parallel operation mode discussed within Section II, a sequential operation mode may be envisaged that exploits in a stronger way the concept of deflation [12]. Strictly speaking, we may suppose the network be formed by a only neuron described by the input–output relationship $y = \mathbf{w}^H\mathbf{x}$, trained by the rule (6). It is able to extract a component at a time. In order to extract all source signals, it is sufficient to present the neuron with input streams deflated from the already extracted components; namely, for extracting kth source signal, we construct the input stream $\mathbf{x}_k$ as

$$\mathbf{x}_1(t) \stackrel{\text{def}}{=} \mathbf{x}(t), \quad \mathbf{x}_{k+1}(t) \stackrel{\text{def}}{=} \mathbf{x}_k(t) - y_k(t)\tilde{\mathbf{w}}_k, \qquad k \geq 1 \quad (7)$$

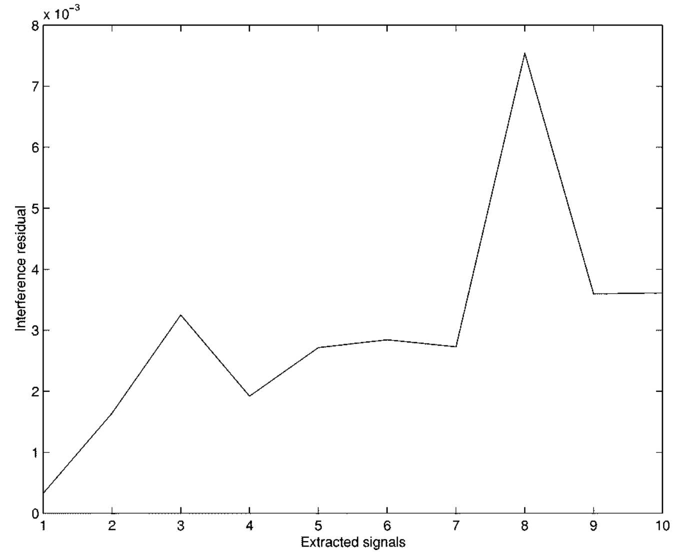

Fig. 4. Interference residuals for ten source signals mixed and successfully separated with the sequential version of EHA algorithm.

where $\tilde{\mathbf{w}}_k$ is an estimation of the "orthogonal signature vector" corresponding to the kth extracted component and $y_k = \mathbf{w}_k^H\mathbf{x}_k$. It may be obtained [12] by an algorithm designed to minimize $E[\|\mathbf{x}_{k+1}\|^2]$, which in the present complex-valued case reads

$$\frac{d\tilde{\mathbf{w}}_k}{dt} = -\tilde{\eta}E[y_k^\star \mathbf{x}_{k+1}] \quad (8)$$

with $\tilde{\eta} > 0$ being a properly tuned learning step-size.

In contrast to the parallel version, which proves to give reliable results and to work in an efficient way with relatively few sources, the sequential (or *cascade*) version generally performs the best on larger scale problems [12].

As an example, let us consider the separation of four QAM4 signals, three PSK8 signals, and three QAM16 signals. The result is expressed as the interference residual (defined as in [12] except for the normalization factor $1/m$) pertaining to each extracted component and is shown in Fig. 4. It has been obtained by setting $\eta = \tilde{\eta} = 0.001$ and by running the proposed algorithm on 20 000 sources' samples for each component. The interference residuals have very good values, as visually confirmed by the Fig. 5, which illustrates the ten source signals as recovered by using the EHA algorithm.

However, the main disadvantage of the sequential approach is that extraction errors inevitably accumulate from stage to stage because, contrary to parallel version, a neuron cannot help other neurons in ameliorating their separation capabilities. This is clearly evidenced by the (slowly) increasing values of errors in the Fig. 4.

C. A Comparison of Computational Complexity

In order to better discuss the advantages and drawbacks of the parallel and sequential operation modes, we also compare two parameters: the number of floating point operations (flops) required for the algorithms to run, and the elapsed CPU times in seconds (on a 450 MHz, 64 MB machine), both averaged over the total number of extracted components and over the total number of learning cycles.

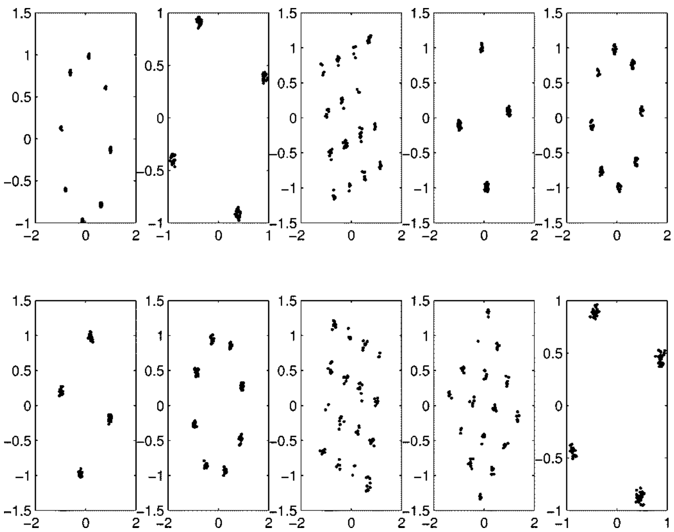

Fig. 5. Neuron output for ten source signals mixed and successfully separated with the sequential version on EHA algorithm. Sources are numbered from one to ten from top/left to bottom/right.

The results for the sequential version are 4123 averaged flops and about $6.5195 \cdot 10^{-4}$ s/cycle/component. For the parallel version we obtained: 566 averaged flops and $1.4906 \cdot 10^{-4}$ s/cycle/component. In conclusion, the parallel version looks more efficient, likely because the deflation operations are inherently exploited rather than explicitly computed.

V. Conclusion

Within this letter, we presented simulation results confirming the effectiveness of the proposed approach, either in the parallel and in the sequential version. However, we noted that parallel version is unable to converge properly with many sources, while the sequential version suffers of progressive output quality degradation. Also, the sequential version cannot be employed in on-line operation, where the parallel version is the only solution. It can thus be envisaged that a hybrid scheme, when applicable, would perform the best. It consists in the parallel extraction of 4/5 sources only, a deflation of the observed data from the extracted components, and the reiteration of the procedure until all sources have been extracted.

The present letter is aimed at giving two main contributions: 1) to extend classical Hebbian learning to complex-valued counterpart (following some interesting closely-related works as, e.g., the one concerning RBF-networks [3]) and to non-quadratic (nonlinear) optimization, and 2) to suggest a way to properly choose the involved nonlinear function on the basis of the Sudjianto–Hassoun interpretation of nonclassical Hebbian learning.

References

[1] E. Bingham and A. Hyvärinen, "ICA of complex valued signals: A fast and robust deflationary algorithm," in *Proc. Int. Conf. Neural Networks*, vol. 3, July 2000, pp. 357–362.

[2] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.

[3] S. Chen, S. McLaughlin, and B. Mulgrew, "Complex-valued radial basis function networks—Part I: Network architecture and learning algorithms," *Signal Process.*, vol. 35, pp. 19–31, 1994.

[4] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[5] A. Hyvärinen and E. Oja, "Independent component analysis by general nonlinear Hebbian-like rules," *Signal Process.*, vol. 64, no. 3, pp. 301–313, 1998.

[6] S. Fiori, "Blind separation of circularly-distributed sources by neural extended APEX algorithm," *Neurocomputing*, vol. 34, pp. 239–252, Aug. 2000.

[7] ——, "A general class of ψ-APEX PCA neural algorithms," *IEEE Trans. Circuits Syst.*, vol. 47, pp. 1394–1398, Sept. 2000.

[8] ——, "Blind signal processing by the adaptive activation function network," *Neural Networks*, vol. 13, pp. 597–611, Aug. 2000.

[9] J. Karhunen and J. Joutsensalo, "Generalizations of PCA, optimization problems, and neural networks," *Neural Networks*, vol. 8, pp. 549–562, 1995.

[10] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks*, vol. 2, pp. 459–473, 1989.

[11] A. Sudjianto and M. H. Hassoun, "Nonlinear Hebbian rule: A statistical interpretation," in *Proc. Int. Conf. Neural Networks*, vol. 2, 1994, pp. 1247–1252.

[12] R. Thawonmas, A. Cichocki, and S.-I. Amari, "A cascade neural network for blind signal extraction without spurious equilibria," *IEICE Trans. Fundamentals*, vol. E-81A, pp. 1833–1846, Sept. 1998.